Naval Dental Research Institute
NDRI-PR 93-02
AUGUST 1993

SIMULATION STUDY OF METHODS TO DETECT PERIODONTAL ASSOCIATIONS WHEN THEY ARE INCONSISTENT AMONG SUBJECTS

M. E. COHEN

Naval Medical Research and Development Command
Bethesda, Maryland

NAVAL DENTAL RESEARCH INSTITUTE
BUILDING 1-H
GREAT LAKES, ILLINOIS 60088-5259

SIMULATION STUDY OF METHODS TO DETECT PERIODONTAL ASSOCIATIONS WHEN THEY ARE INCONSISTENT AMONG SUBJECTS

M. E. COHEN

Research Progress Report NDRI-PR 93-02
Work Unit M0095.003-0003
Naval Medical Research and Development Command
National Naval Medical Center
8901 Wisconsin Avenue
Bethesda, Maryland 20889-5606

Approved and released by:

S. A. RALLS
Captain, Dental Corps
United States Navy
Commanding Officer

Simulation study of methods to detect periodontal associations when they are inconsistent among subjects

M. E. Cohen
Naval Dental Research Institute, Great Lakes, Illinois, USA

Cohen ME: Simulation study of methods to detect periodontal associations when they are inconsistent among subjects. Community Dent Oral Epidemiol 1993; 21: 19-23.

Abstract – Most statistical methods used to evaluate associations between indices of clinical periodontal diseases and purported prognostic markers test for effects across subjects. If associations exist within only a subset of subjects, however, associations may be masked, particularly in small studies. This issue was explored by using simulation to study four methods for detecting periodontal associations. Built into the simulations was the possible biological reality that a non-zero association between the two variables of interest (squared correlation coefficients, ρ^2, ranged from 0.1 to 0.9 depending on simulation), measured at 16 sites per subject, did not exist in all of 10 hypothetical subjects. The four methods for testing the null hypothesis that $\rho = 0$, or a related hypothesis; were: (1) *Sites*, analysis based on 160 sites incorrectly considered independent observations; (2) *Subjects*, analysis based on one score for each of 10 subjects; (3) *Each subject*, separate analyses based on sites within each of 10 subjects, family-wise type I (α) error corrected for multiplicity, and (4) the *Each Subject* method where P-levels were estimated using permutation procedures rather than t-distributions. *Each Subject* methods were found to have greater relative power (although there are differences in null hypotheses) under conditions of heterogeneity in ρ and are considered to be particularly relevant in exploratory periodontal research when the primary interest is establishing the existence of a relationship, even if in only a subset of subjects.

Key words: biostatistics, periodontal associations

Naval Dental Research Institute, Bldg 1-H, Great Lakes, Illinois 60088-5259, USA

Accepted for publication 1 May 1992

Periodontal research is sometimes directed towards estimating the degree of relationship between the primary clinical variable of attachment loss and potential markers of current or future disease activity such as changes in gingival crevicular fluid (GCF) contents and quantities of various microorganisms. Researchers frequently study multiple sites over time in more than one patient, and are interested in testing hypotheses about population parameters.

In two recent studies (1, 2), for example, attachment loss was related to the enzyme aspartate aminotransferase (AST) in GCF. Data were analyzed at many levels, including patients, teeth within patients, and sites within teeth, but tests were directed towards population parameters. This is appropriate since AST, being associated with cell death, should have a consistent relationship to periodontal destruction in all sites in all subjects.

In the case of other possible periodontal predictors, however, such as the presence of a specific microorganism, the hypothesis of a universal and consistent association is not as compelling. The biological reality of the situation may be, for example, that attachment loss is related to different microorganisms in different subjects at different times. The presence of such effect heterogeneity is not unusual in epidemiological research and is typically compensated for by conducting studies with large sample sizes. In research on periodontal microbiota, however, determination of "risk exposure" and evaluation of "disease status" are expensive and labor intensive so that studies which are "large enough" may not be routinely feasible.

Given that periodontal studies of this nature may therefore be chronically underpowered relative to rejecting the traditional null hypothesis that association within the population is zero, a single subject strategy (3) may be useful for exploratory purposes. This strategy does not address the clearly more vital issue of association in the population, but may provide preliminary information that could both justify further research and make subsequent studies more efficient. The present research evaluates four approaches to data analysis, including testing of single subjects, under the biological assumption that the association between a disease and a selected marker exists only in some subjects tested.

Method

Simulated trials (see Appendix A for consideration of the analytic alternative)

were constructed where hypothesized attachment levels and counts of a specific microorganism were sampled at 16 sites in each of 10 patients on two occasions separated by time. Interest is directed towards the calculated correlation coefficient (r) between change in microorganism counts (X) and change in attachment level (Y). A biological reality constructed into the simulations was that the microorganism was associated (population correlation coefficient, ρ, is greater than zero) in only one through 10 subjects in the sample. For computational simplicity in the simulations, changes at the 16 sites ($i = 1,\ldots,16$) in X and Y were random bivariate standard normal deviates. Four approaches for testing statistical significance were studied. In all cases, tests were two-tailed (critical r-values refer to absolute magnitude) with type I (α) error set at 0.05.

Sites - The 160 sites (16 in each of 10 subjects) were incorrectly considered as independent observations and the null hypothesis that $\rho = 0$ was tested against the alternative hypothesis that $\rho \neq 0$. With $n = 160$, critical $r = 0.15523$. Since sites are actually not independent, the validity of this method for making inferences to the population is compromised.

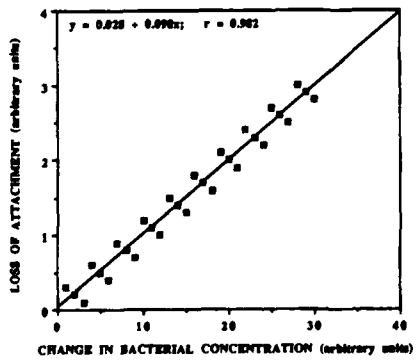

Fig. 1. To demonstrate that the correlation computed on the basis of all sites may differ from that computed within any or all subjects, the figure shows hypothetical data for three sites in 10 subjects. The correlation is -1 for sites within every subject but close to $+1$ across all sites, although the latter analysis is not valid for inferential purposes. If the correlation was computed on mean subject data so that the independence problem is eliminated, the r-value would be $+1$. Such divergence in biological effects estimated within subjects versus across subjects would not be routinely anticipated but nevertheless highlights that *Sites* analysis addresses a conceptually different association than that of *Subjects* or *Each Subject*, irrespective of the validity issue.

However, this approach is unfortunately common in the literature (4) and is included for comparative purposes.

In addition to inferential problems associated with site independence, this approach makes evident a related issue in that the *Sites* r-value need not be similar to any r observed among sites within individual subjects. For example, the within subject r between microbial concentration and attachment loss may be negative within every subject but because of subjects' data location on a scale, positive across all *Sites*. This is depicted in Fig. 1. The relationship inherent in the *Sites* r-value can be validly estimated by computing a correlation between microbial concentration and attachment, based on means (across sites) for each subject. This, however, does not address site specificity in the same way as when sites within subjects are studied.

Subjects - The r for each of the 10 subjects was computed separately and the null hypothesis that mean $\rho = 0$ was tested by a one-sample t-test. With $n = 10$ (10 by subject r-values), critical $t = 2.26216$. This method treats each subject's r as an index score representing a summary characteristic.

Each Subject (parametric) - An r was computed for each of the ten subjects and each was tested for statistical significance. With $n = 16$ (pairs of observations within each subject), critical $r = 0.49731$. This significance test does not allow for inferences about the population (except that $\rho = 0$ for every subject in the population) but can be used to identify whether an effect exists in a particular subject. (Appendix B discusses another analysis strategy which uses individual P-values to test an overall null hypothesis.)

Since 10 subjects are tested in each trial the probability of at least one subject having a significant r-value when each is tested at $\alpha = 0.05$, and all nulls are true, is not 0.05, but rather $1 - 0.95^{10} = 0.40126$. This can be conceptualized as a family-wise α-error rate associated with the family-wise null hypothesis that $\rho = 0$ for every subject in the set.

Several methods have been proposed to adjust P-levels used for significance testing so that the family-wise α-error rate, rather than the individual rate, is maintained at 0.05. Consider the set of m null hypotheses $H_{(1)},\ldots,H_{(m)}$ and the P-values, $P_{(1)},\ldots,P_{(m)}$, corresponding to these hypotheses. A Bonferroni correction would require that $P_{(i)} < \alpha/m$ for rejection of each $H_{(i)}$. This approach is overly conservative. More accurate procedures which maintain the family-wise α-error rate at 0.05, but have increased power to detect effects when they exist in individual subjects, have been proposed by HOLM (5) and by HOCHBERG (6). The latter method will always be as good or better than the former (7) and was used in these simulations.

HOCHBERG's method requires that P-values are first ordered from smallest or most significant ($i = 1$) to largest or least significant ($i = m$) and that the $H_{(i)}$s are then sequentially tested in reverse order. Each of the $H_{(i)}$s is retained so long as $P_{(i)} > \alpha/(m - i + 1)$. The first time that the inequality is reversed, that $H_{(i)}$ and all remaining $H_{(i)}$s are rejected. In the present simulation where $m = 10$, if the largest P-value, $P_{(10)}$, is less than $0.05 = (0.05/(10 - 10 + 1)$, both tails) then all $H_{(i)}$s are rejected. If this is not the case, then if the second largest P-value, $P_{(9)}$ is less than $0.025 = (0.05/(10 - 9 + 1))$ then the remaining nine $H_{(i)}$s are rejected, and so forth. To reduce computations, critical r-values rather than P-values were used in the simulations. The critical absolute r-values (based on $n = 16$) corresponding to P-values (in one-tail) of 0.02500, 0.01250, 0.00833, 0.00625, 0.00500, 0.00417, 0.00357, 0.00313, 0.00278, and 0.00250 were 0.49731, 0.55702, 0.58771, 0.60782, 0.62259, 0.63411, 0.64362, 0.65145, 0.65832, and 0.66434, respectively. The family-wise null hypothesis was considered rejected if any of $H_{(i)}$ individual null hypotheses were rejected.

Power for the first three test methodologies: *Sites*, *Subjects*, *Each Subject* (parametric), was evaluated by conducting 10000 trials for each of 90 conditions formed by the factorial combination of 9 levels of correlation ($\rho^2 = 0.1, 0.2, \ldots, 1.0$) and 10 levels of number of subjects exhibiting that association (1 to 10). Other simulations, which are not reported, verified that alpha error levels were as expected.

The correctness of P-values used in these simulations is dependent upon the assumption (among others) that each measurement represents an independent observation. For *Sites* and *Each Subject* (parametric) methodologies, this is not the case. For the latter method, the problem of appropriate P-values can be avoided by estimating them in terms of

a random assignment model rather than a random sampling model. This can be accomplished by using permutation tests.

Each Subject (permutation) – The r is first computed for the subject's original set of 16 pairs of observations. Data array location (i.e., position in the 16 "slots") for one of the variables, say Y, is then permutated by a random process to Y*, so that pairings between x_i and y_i* themselves become random. Specifically, a new ordering of Y is generated such that y_1* is randomly selected out of the 16 original scores, y_2* is randomly selected out of the remaining 15 scores, and so forth. This process is repeated many times, say 9999, and an r computed each time. The P-value corresponding to the original r-value is then represented by the proportion of the 10000 data sets where r is greater than or equal to the original r.

The permutation P-value is not dependent upon parametric or random sampling assumptions and is appropriate whenever single subject designs are utilized. However, generation of permutation statistics is computationally intensive. This is not a limitation in clinical research where a personal computer can produce a P-value in a few seconds, but does pose a problem for a simulation study where thousands of such P-values are required. At 9999 permutations for each of 10 subjects for each of 90 simulation sets of 10000 trials each, the present study would require an additional 89999100000 computations of r.

In order to evaluate how closely the permutation P-value is approximated by the parametric P-value, 1000 trials were simulated for each of four of the 90 conditions previously described. The four conditions were for $\rho^2 = 0.3$ for the case where one, two, three, or four subjects exhibited the relationship. Selection of these conditions was done after the initial findings were produced so that comparisons for a reasonable range of power levels could be evaluated. All procedures were the same as in the prior simulations except that additionally, the P-value for each subject's r was estimated on the basis of 10000 data permutations (including the original data set).

Results

Tables 1 through 4 show power when 1 to 10 subjects had associations between X and Y resulting from random sampling of bivariate standard normal distributions with ρ^2 values of 0.1, 0.3, 0.5 and 0.7. These values correspond to low, moderate, high, and very high correlations, respectively ($\rho = 0.316$, 0.548, 0.707, and 0.837). Simulations that were previously described for other values of ρ^2 are not reported here since they are consistent with the trends established in these tables.

Analysis by *Sites* is always more powerful than analysis by *Subjects* under the conditions of these simulations. This would seem to be the motivation for the appearance of site analysis in some research where problems associated with the lack of independence have been ignored. Analysis by *Subjects* tended to be more powerful than analysis by *Each Subject* when (non-zero) ρ values were low and present in many subjects. Thus, analysis by *Subjects* showed greater relative power under conditions of greatest homogeneity. When there was heterogeneity in associations, the *Each Subject* method was more successful is achieving statistical significance. In the extreme, when $\rho^2 = 0.7$ for one or two subjects, analysis by *Subjects* had power in the vicinity of the alpha error level. Apparently, power was depressed by heterogeneity contributing to error variance. In contrast, analysis by *Each Subject* had power exceeding 0.95. Logically, this is not a remarkable finding but underscores the improvement in power which is possible when a single subject approach is adopted under these conditions.

Table 1. Percentage of 10000 trials where a null hypothesis was correctly rejected (power × 100). $\rho^2 = 0.1$ for the number of subjects indicated. $\rho^2 = 0.0$ for the remainder of the 10 subjects

Subjects	Sites[a]	Subjects	Each subject[b]
1	7.04	5.13	8.79
2	12.90	7.64	12.78
3	22.38	11.95	16.80
4	36.01	18.73	20.30
5	52.10	29.08	24.10
6	67.22	41.15	26.93
7	80.79	56.33	30.05
8	90.15	71.06	31.91
9	95.59	84.70	35.51
10	98.42	93.52	38.83

[a] As noted in the text, analysis by sites is not valid. Power is reported for limited comparative purposes and should not be considered as legitimate.
[b] Power here refers to the event that the null hypothesis was rejected for at least one subject.

Table 2. Percentage of 10000 trials where a null hypothesis was correctly rejected (power × 100). $\rho^2 = 0.3$ for the number of subjects indicated. $\rho^2 = 0.0$ for the remainder of the 10 subjects

Subjects	Sites[a]	Subjects	Each subject[b]
1	11.24	4.85	30.22
2	28.60	9.68	49.05
3	55.23	18.78	62.54
4	79.49	33.63	72.55
5	93.80	54.94	79.92
6	98.79	77.44	85.67
7	99.83	92.47	88.95
8	100.00	98.87	92.24
9	100.00	99.95	93.98
10	100.00	100.00	95.80

[a,b] See Table 1.

Table 3. Percentage of 10000 trials where a null hypothesis was correctly rejected (power × 100). $\rho^2 = 0.5$ for the number of subjects indicated. $\rho^2 = 0.0$ for the remainder of the 10 subjects

Subjects	Sites[a]	Subjects	Each subject[b]
1	15.00	4.46	66.71
2	42.95	9.99	89.20
3	76.60	20.71	95.97
4	94.46	41.42	98.70
5	99.35	69.19	99.53
6	99.94	90.51	99.83
7	100.00	99.00	99.97
8	100.00	99.98	99.96
9	100.00	100.00	99.98
10	100.00	100.00	100.00

[a,b] See Table 1.

Table 4. Percentage of 10000 trials where a null hypothesis was correctly rejected (power × 100). $\rho^2 = 0.7$ for the number of subjects indicated. $\rho^2 = 0.0$ for the remainder of the 10 subjects

Subjects	Sites[a]	Subjects	Each subject[b]
1	19.11	3.87	95.16
2	56.04	8.73	99.70
3	88.08	22.47	99.99
4	98.66	47.08	100.00
5	99.82	77.56	100.00
6	100.00	95.74	100.00
7	100.00	99.88	100.00
8	100.00	100.00	100.00
9	100.00	100.00	100.00
10	100.00	100.00	100.00

[a,b] See Table 1.

Permutation procedures exhibited power that was very close to parametric alternatives. At $\rho^2 = 0.3$ for one to four subjects, *Each Subject* (parametric) simulations based on 1000 trials found powers of 0.303, 0.485, 0.667, and 0.745, respectively. When P-levels were based on permutation procedures powers of 0.308, 0.487, 0.651, and 0.734 were observed.

Discussion

The results show that rejection of the traditional null hypothesis that mean $\rho = 0$ is infrequent in small studies when the effect of interest exists in only a small subset of subjects. This lack of power could be one reason why the incorrect analysis of sites as independent observations has been common. However, both approaches were less powerful than *Each Subject* analysis under conditions of heterogeneity.

Analysis using *Each Subject* (parametric and permutation) approaches however, clearly restricts inference. It does not address the null hypothesis that mean ρ in the population is zero. But, if it is true that periodontal destruction is associated with different microorganisms in different subjects or in the same subject at different points in time, then appropriate testing of the traditional null hypothesis may require sample sizes that are not routinely available (although meta analysis may be useful in addressing this problem).

Single subject designs are not common in the biomedical literature, though there are some exceptions (e.g., 8, 9). It appears that they are most valuable when designed to construct a patient specific treatment regime (8) or to demonstrate the simple existence of rare phenomena (9). In periodontal research, findings of statistical significance at the level of individual subjects may serve to justify and improve the efficiency of more definitive research.

The periodontal literature contains many studies in which association of disease with purported microbiological markers has not been confirmed. Recently, for example, LISTGARTEN *et al.* (10) report "... the results indicate that the presence of the above bacterial species cannot of itself serve as a reliable predictor of future episodes of recurrent disease ..." and MAGNUSSON *et al.* (11) report "... few, if any, of the "classical" pathogens were detected in the plaque samples obtained at the time progressive disease was diagnosed." Absence of effects have, at times, been attributed to methodological limitations, justifying the development and use of more sensitive devices and procedures, and collection of data from larger samples. Analyses on individual subjects may be useful in evaluating such recommendations.

Consider that the across-subjects correlation between counts of a particular microorganism and change in attachment level is small and not statistically significant, but that an association has been identified in a few subjects using procedures to control for family-wise α-error level. At this point, it is known (at $P < 0.05$) that an association exists in some subjects in the population. This justifies additional research to estimate the prevalence of such associations and to develop procedures to identify subjects that are of this associative type. Further research directed specifically at these subjects might be more efficient. In contrast, if findings for individual subjects are also negative, continued research seems less justified.

The *Each Subject* approach can be considered to lie between purely exploratory versus confirmatory data analysis. The former is most concerned with revealing patterns and features of the data and the latter with the reproducibility of those patterns and features in terms of statistical significance (12). In exploratory analysis correlations might be examined and if there is evidence of heterogeneity, groups would be created based on biological explanations. However, if the biological bases for heterogeneity are not well founded then the analyst runs a risk of creating an "effect" solely on the basis of post hoc subject selection. In confirmatory analysis, consistency of effect may not explored. If consistency is absent, however, then the analysis lacks power. The *Each Subject* approach is exploratory in the sense that it is directed towards effects which are heterogeneously distributed in the population but confirmatory in the sense that it controls for the multiplicity problem inherent in hypothesis testing of many subjects.

It can be argued that the *Each Subject* approach, by being potentially sensitive to an association in only a single subject, may provide findings that are not important to understanding the disease process in general. This is a valid and well founded epidemiological criticism but it is more relevant to the present progress and direction of periodontal research than to the statistical methodology. There may be very many periodontal pathogens and the biological basis may not be available for identifying one microorganism at a site as "the" pathogen versus all other microorganisms at that site. Therefore, the post hoc microbiological grouping of patients may be difficult to justify and the alpha error control inherent in the *Each Subject* methodology important.

It seems that three analytic strategies may be selected based on a preliminary screening of the data. If associations are homogeneous then a *Subjects* type analysis should be used. If associations are heterogeneous but there is a reasonable biological basis for post hoc grouping then methods should be used which assume homogeneous associations within groups and heterogeneous associations between groups. If justifications for grouping are not forthcoming then the *Each Subject* approach may provide a valid test of the hypothesis that there is an association in some subjects.

In (unreported) simulations, α-error using parametric P-values was at nominal levels. Therefore, the use of permutation P-levels, which in simulation was shown to have similar power, would not appear to be required. However, measurements in the simulations were constructed to be bivariate standard normal deviates. In actual research settings nonnormality is common and, therefore, the use of parametric P-values may be potentially unacceptable and the use of permutation P-levels necessary.

The basic requirement of single subject designs is that there must be repeated observations within a subject, either over time or space, that provide a context for the randomization (to groups or to pairs). Periodontal sites within mouths would seem to offer numerous single subject or site test possibilities.

Although the issue of analysis at the level of the individual site has not been addressed, it is clear that the principles of permutation tests apply to them as well as to individual subjects. For example, an r might be computed between two variables observed in the the same site on 12 occasions over a three year period. If it is true that different microorganisms

are pathogenic at different sites, then it is logical to test for associations separately at different sites. The only difficulty arises if one is studying very many sites, say 16 in each of 10 subjects. With 160 "experiments" significant P-values may have to be as small as 0.0003 (0.05/160). This problem might be overcome by an *a priori* selection of sites based on risk.

Acknowledgments - The author recognizes many insightful suggestions made by the manuscript reviewers. The opinions expressed herein are those of the author and cannot be construed as reflecting the views of the Navy Department or the Naval Service at large. The use of commercially available products does not imply endorsement of these products or preference to other similar products on the market. Supported by Naval Medical Research and Development Command Project Number M0095.003-0003.

References

1. Chambers DA, Imrey PB, Cohen RL, Crawford JM, Alves MEAF, McSwiggin TA. A longitudinal study of aspartate aminotransferase in human gingival crevicular fluid. *J Periodont Res* 1991; *26*: 65–74.
2. Imrey PB, Crawford JM, Cohen RL, Alves MEAF, McSwiggin TA, Chambers DA. A cross-sectional analysis of aspartate aminotransferase in human gingival crevicular fluid. *J Periodont Res* 1991; *26*: 75–84.
3. Edgington ES. *Randomization tests*. 2nd edn. New York: Marcel Dekker, 1987: 244–80.
4. Emrich LJ. Common problems with statistical aspects of periodontal papers. *J Periodontol* 1990; *61*: 206–8.
5. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat* 1979; *6*: 65–70.
6. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 1988; *74*: 800–3.
7. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med* 1990; *9*: 811–8.
8. McLeod RS, Taylor DW, Cohen A, Cullen JB. Single patient randomised clinical trial: Its use in determining optimal treatment for patient with inflammation of a Koch continent ileostomy reservoir. *Lancet* 1986; *I*: 726–8.
9. Weiss B, Williams JH, Margen S, Abrams B, Caan B, Citron LJ, Cox C, McKibben J, Ogar D, Schultz S. Behavioral responses to artificial food colors. *Science* 1980; *287*: 1487–9.
10. Listgarten MA, Slot J, Nowotny AH, Oler J, Rosenberg J, Gregor B, Sullivan P. Incidence of periodontitis recurrence in treated patients with and without cultivable *Actinobacillus actinomycetemcomitans*, *Prevotella intermedia*, and *Porphyromonas gingivalis*: a prospective study. *J Periodontol* 1991; *62*: 377–86.
11. Magnusson I, Marks RG, Clark WB, Walker CB, Low SB, McArthur WP. Clinical, microbiological and immunological characteristics of subjects with "refractory" periodontal disease. *J Clin Periodontal* 1991; *18*: 291–9.
12. Kotz S, Johnson NL (Editors-in-Chief). *Encyclopedia of statistical sciences*, vol. 2. New York: Wiley, 1982: 579.
13. Kotz S, Johnson NL (Editors-in-Chief). *Encyclopedia of statistical sciences*, vol. 9. New York: Wiley, 1988: 222–3.

Appendix

A. Simulation was chosen to estimate power since it would be sufficiently accurate for these purposes and could easily be applied to all procedures studied. However, as suggested by manuscript reviewers, it would seem that results by analysis could provide more mathematically elegant solutions. Such analytical solutions, though, would be different for each methodology and there are indications that some of these solutions would be more difficult to achieve than others.

B. The *Subjects* methodology addresses the null hypothesis that mean ρ across subjects is zero, while the *Each Subjects* approach tests the null that $\rho = 0$ for each subject. A third alternative, not studied here, tests, through procedures where individual P-values are combined, the null hypothesis that all within subject nulls are true versus the alternate hypothesis that the null is false for at least one subject. This alternative is required when data from the different "experiments" (or subjects) cannot be reasonably pooled (e.g., where different parameters are being studied in different experiments) so that the *Subjects* methodology is precluded. The methodology may also be used as an alternative to weighting when there are unequal number of sites per subject, since P-values will incorporate weighting automatically. It is not, however, a reasonable alternative to the *Each Subjects* methodology because it does not identify the specific subjects for which the null is not true.

When the combined P-value approach is advisable, Westberg (13) notes that Fisher's procedure, based on the product of P-values, is most powerful when many individual nulls are false to a comparable degree and Tippett's procedure, based on the minimum P-value is most powerful when one null is false and very deviant. Westberg also described her own "adaptive method" that may have the advantages of both procedures.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* SIMULATION STUDY OF METHODS TO DETECT PERIODONTAL ASSOCIATIONS WHEN THEY ARE INCONSISTENT AMONG SUBJECTS		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER NDRI-PR 93-02
7. AUTHOR(s) M. E. COHEN		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Dental Research Institute Building 1-H Great Lakes, IL 60088-5259		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS M0095.003-0003
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Medical Research and Development Command National Naval Medical Center, 8901 Wisconsin Ave. Bethesda, MD 20889-5606		12. REPORT DATE August 1993 13. NUMBER OF PAGES (5)
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Chief, Bureau of Medicine and Surgery Navy Department 2300 E Street NW Washington, DC 20372-5300		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

This document has been approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

This document has been approved for public release; distribution unlimited.

18. SUPPLEMENTARY NOTES

Community Dentistry and Oral Epidemiology 1993; 21: 19-23

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Biostatistics, Periodontal Associations

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

Most statistical methods used to evaluate associations between indices of clinical periodontal diseases and purported prognostic markers test for effects across subjects. If associations exist within only a subset of subjects, however, associations may be masked, particularly in small studies. This issue was explored by using simulation to study four methods for detecting periodontal associations. Built into the simulations

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

was the possible biological reality that a non-zero association between the two variables of interest (squared correlation coefficients, ρ^2, ranged from 0.1 to 0.9 depending on simulation), measured at 16 sites per subject, did not exist in all of 10 hypothetical subjects. The four methods for testing the null hypothesis that $\rho=0$, or a related hypothesis; were: (1) *Sites*, analysis based on 160 sites incorrectly considered independent observations; (2) *Subjects*, analysis based on one score for each 10 subjects; (3) *Each subject*, separate analyses based on sites within each of 10 subjects, family-wise type I (α) error corrected for multiplicity, and (4) the *Each Subject* method where P-levels were estimated using permutation procedures rather than t-distributions. *Each Subject* methods were found to have greater relative power (although there were differences in null hypotheses) under conditions of heterogeneity in ρ and are considered to be particularly relevant in exploratory periodontal research when the primary interest is establishing the existence of a relationship, even if in only a subset of subjects.